Journal of Educational Psychology

Models as Feedback: Developing Representational Competence in Chemistry

Shamin Padalkar and Mary Hegarty

Online First Publication, August 11, 2014. http://dx.doi.org/10.1037/a0037516

Journal of Educational Psychology

0022-0663/14/$12.00 http://dx.doi.org/10.1037/a0037516

Models as Feedback: Developing Representational Competence in Chemistry

Shamin Padalkar and Mary Hegarty
University of California, Santa Barbara

Spatial information in science is often expressed through representations such as diagrams and models. Learning the strengths and limitations of these representations and how to relate them are important aspects of developing scientific understanding, referred to as *representational competence*. Diagram translation is particularly challenging for students in organic chemistry, and although concrete models greatly help in solving diagram translation problems, most students do not use models spontaneously. In 2 experiments, we examined the effectiveness of instructional interventions for teaching diagram translation using models. In Experiment 1, students drew diagrams and checked their accuracy by attempting to match concrete models to their solutions (model-based feedback). The instruction helped students in the experimental group to identify their mistakes, understand the usefulness of concrete models, and led to large improvements in performance, compared with a control group. To examine whether feedback, the opportunity to match models, or both was the critical aspect of the intervention, in Experiment 2, 1 group was provided only verbal feedback (by a tutor) and another group matched diagrams and concrete models, but not in the context of receiving an evaluation of their pretest performance. Feedback alone did not improve performance relative to a control group, but the opportunity to match models and diagrams improved performance relative to control. The results indicate that using models as feedback is an effective way of training representational competence in the domain of organic chemistry and more generally in science, technology, engineering, and mathematics disciplines.

Keywords: spatial tools, feedback, chemistry education

Spatial information, such as shape, size, structure, and motion, is particularly important in the natural sciences, and certain branches of science are devoted to studying spatial properties. For example, anatomy is the study of the structure of living things, geology is the science of the structure of the earth, and stereochemistry is the study of the structure of compounds. Understanding spatial information is particularly challenging when the relevant structures are not directly observable, for example, because they occur at a scale of space that is not visible (e.g., molecules) or are internal to some three-dimensional (3-D) structure that we typically only see from the outside (e.g., internal anatomy). In these cases, spatial information is represented most directly through spatial representations such as diagrams, concrete and virtual models, and animations; it is also represented using non-spatial representations such as text, symbols, and formulae.

Shamin Padalkar and Mary Hegarty, Department of Psychological and Brain Sciences, University of California, Santa Barbara.

This research was supported by National Science Foundation Grant 1008650. We are thankful to Andrew Stull, Trevor Barrett, David Sanosa, Emily Steiner, Mike Stieff, and Bonnie Dixon for their contributions to the research.

Correspondence concerning this article should be addressed to either Shamin Padalkar, who is now at the Inter-University Centre for Astronomy and Astrophysics (IUCAA), Post Bag 4, Ganeshkhind, Pune University Campus, Pune, Maharashtra 411007, India, or Mary Hegarty, Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106. E-mail: pshamin@gmail.com or mary.hegarty@psych.ucsb.edu

Scientists are facile in using a range of different representations of spatial phenomena that vary in their dimensionality (e.g., two-dimensional [2-D] vs. 3-D), abstraction (e.g., diagrams vs. equations), and spatial perspective (e.g., orthographic vs. isometric projections, cross-sections) to represent, reason, and communicate about different spatial phenomena (Ainsworth, 2006; Kozma & Russell, 1997, 2005; Lemke, 1998). However, understanding spatial structures from representations has proved to be difficult for students in a range of domains, including anatomy, astronomy, geology, and chemistry (e.g., Chariker, Naaz, & Pani, 2011; Kali & Orion, 1996; Keig & Rubba, 1993; Padalkar & Ramadas, 2010). In these domains, students typically have to master several different spatial representations in introductory classes. Effective use of these multiple representations includes understanding their conventions, how to construct each representation, how to relate alternative representations of the same spatial information, and when to use each representation (Ainsworth, 2006). Kozma and Russell (1997, p. 963) referred to this type of understanding as "representational competence." More reflective and metacognitive uses of multiple representations, including choosing the optimal external representation for a task and inventing new representations, if necessary, has been referred to as "meta-representational competence" (diSessa, 2004, p. 293).

In this article, we focus on organic chemistry, which is a particularly representation-rich domain, and one in which understanding the relevant spatial information (the structure of molecules) is important but complex. Our informal review of a popular organic chemistry textbook (Bruice, 2010) revealed that undergraduate students taking their first organic chemistry course are exposed to at least 10 different kinds of spatial representations of organic molecules. Each representation is specialized for a different purpose in the domain of organic chemistry. Students struggle with mastering these different representations and relating them to each other (Hinze et al., 2013; Keig & Rubba, 1993; Kozma & Russell, 1997; Stull, Hegarty, Dixon, & Stieff, 2012). Here, we report and compare instructional interventions that aim to improve students' representational competence and meta-representational competence in this domain.

Spatial Representations

Scientists use two main types of spatial representations of 3-D spatial structures: models and diagrams. Models, such as the ball-and-stick model of a molecule shown in Figure 1a and desktop models of the solar system in astronomy, represent the 3-D spatial relationships between parts of the referent structure directly. Diagrams, in contrast, represent three dimensions in the two dimensions of the printed page, so typically show a specific perspective or projection of the 3-D entity, and often use conventions to represent three dimensions in two dimensions. Students must remember and interpret these conventions to construct an internal representation of the 3-D structure from the 2-D external diagram. Furthermore, diagrams are static, so inferences about the results of rotations and other spatial transformations involve mentally simulating the transformations, a process that is demanding of spatial

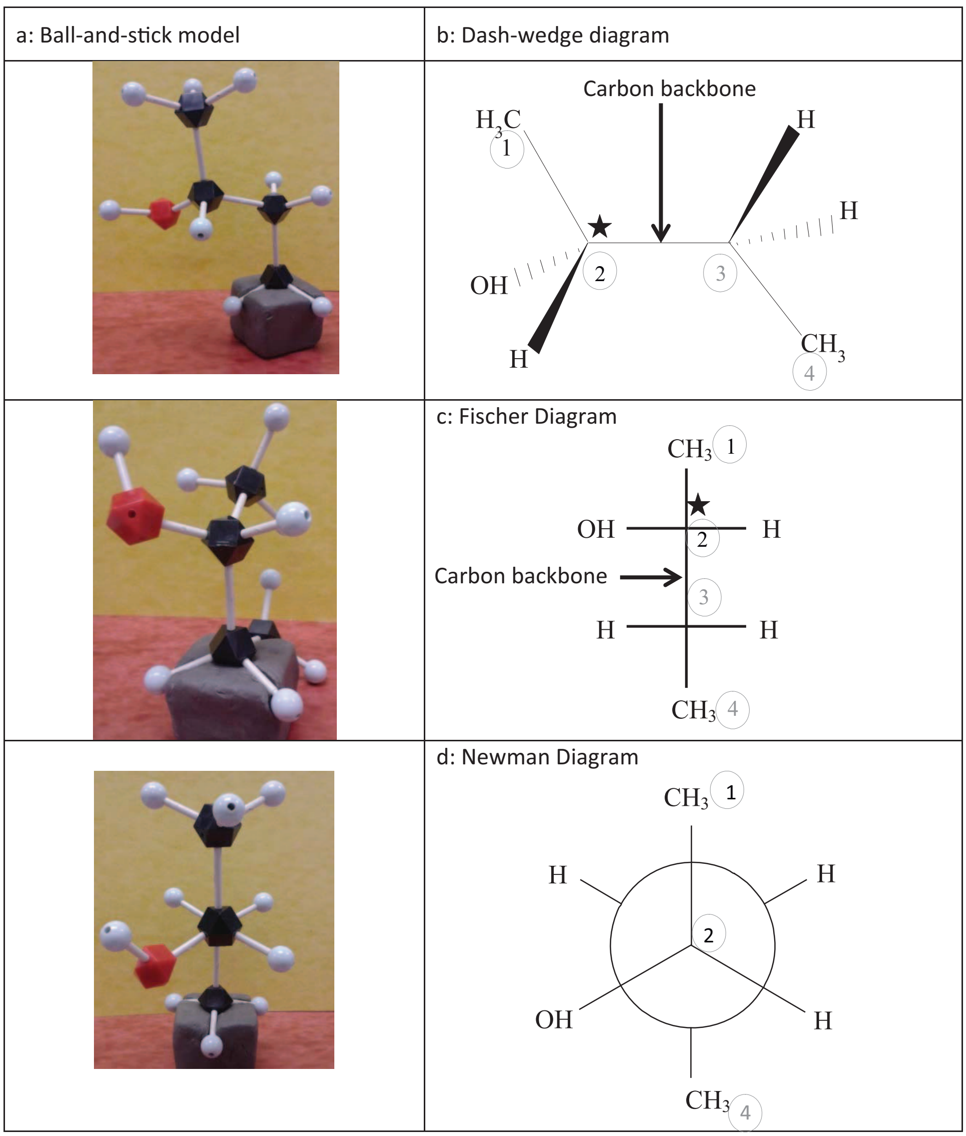


Figure 1. Example of a ball-and-stick model of a molecule, (S)-2-butanol, and Dash-wedge, Fischer, and Newman diagrams of the same molecule. The three different views of the model illustrate the perspective shown in the corresponding diagram. The carbon atoms at Locations 1, 2, 3, and 4 define the carbon backbone. In the diagrams, the carbons at Positions 2 and 3 are not shown explicitly, by convention. In the Newman diagram, the carbon at Position 3 is occluded by the carbon at Position 2. See the online article for a color version of this figure.

working memory (Hegarty, 2004). In contrast, 3-D models often have moving parts, so that spatial transformations (e.g., the rotation of planets or molecular components) can be carried out externally on models. This reduces the demands on spatial working memory and enables students to observe the relevant spatial processes.

Our research addresses four related challenges that students face in introductory science classes. First, although students are introduced to the conventions of different representations in these classes, they often have difficulty translating between representations (Chariker, Naaz, & Pani, 2011; Hinze et al., 2013; Kali & Orion, 1996; Keig & Rubba, 1993; Kozma & Russell, 1997; Novick & Catley, 2007). Second, students have particular difficulty understanding 2-D representations of 3-D structures, such as cross-sections in anatomy and geology (Chariker et al., 2011; Kali & Orion, 1996) or projections in geography (Downs & Liben, 1991). Third, although physical and virtual models can be helpful in teaching students about 3-D structures and scaffolding the understanding of 2-D representations, when provided with models (either physical or virtual), students do not use them effectively, and often ignore them (Keehner et al., 2008; Stull et al., 2012). Merely providing models is not necessarily effective, and researchers have called for the development of pedagogical methods for how to best use models in instruction (e.g., Martin & Schwartz, 2005; McNeil & Uttal, 2009). Fourth, students have poor attitudes toward using models, which stems in part from a lack of understanding of the ontological status of models in science. Specifically, students often perceive models as copies of scientific phenomena rather than as tools for understanding these phenomena or solving problems (Treagust & Chittleborough, 2001).

Representations in Chemistry

To understand how these problems are instantiated in learning organic chemistry, we must first introduce some basic chemistry concepts and terms. Carbon is an essential element in organic compounds. A series of bonded carbon atoms creates a continuous chain known as a *carbon backbone* (indicated in Figure 1b) to which other atoms and groups of atoms (referred to as *substituents*) are bonded. The rotation of substituents (such as OH or NH_2) around a bond results in different *conformations* of a molecule, but does not change the identity or the chemical properties of the molecule.[1] In contrast, molecules made up of the same substituents and connectivity between these substituents, but with different 3-D ordering of the substituents around the carbon backbone, have different chemical properties and are called *stereoisomers* (see Figure 2c). For example, the drug Thalidomide was introduced in late 1950s to treat morning sickness. However, although one isomer of Thalidomide is a cure for morning sickness, the other causes birth defects, which, tragically, resulted in more than 10,000 children being born with deformities. This example illustrates how the spatial arrangement of atoms or substituents in three dimensions can be crucial to the reactivity of a compound.

In the present research, we focus on three diagrammatic representations of the 3-D structure of molecules that students have to master in introductory chemistry classes: Dash-wedge diagrams, Newman projections, and Fischer projections. These three diagrams are informationally equivalent (cf. Larkin & Simon, 1987), but they depict the molecule from three different perspectives and follow different conventions to represent 3-D relations in the two dimensions of the printed page. In Dash-wedge diagrams, the molecule is oriented horizontally (side view), dashed lines represent bonds going into the page, wedged lines represent bonds coming out of the page, and the solid lines represent bonds in the plane of the page (see Figure 1b). In Fischer projections, the two central backbone carbons are oriented vertically in the plane of the paper. The horizontal lines are coming out of the page, and the top and bottom of the vertical line represent bonds are going into the page (see Figure 1c). In Newman projections, the molecule is oriented horizontally with one backbone carbon in front of the other. The front carbon is at the intersection of the three lines, and the rear carbon is behind the circle (see Figure 1d). The Appendix explains the conventions of these three representations in more detail.

Translating between these diagram formats is a common task for chemists and a task that students often have to perform in organic chemistry examinations (American Chemical Society, 2010). It is a good indicator of students' understanding of both the 3-D structure of organic molecules and the conventions of the diagrams. However, it is particularly challenging for students (Keig & Rubba, 1993; Kozma & Russell, 1997). Alarmingly, in an earlier study, Stull et al. (2012) found that students who had been introduced to Dash-wedge, Newman, and Fischer diagrams in an introductory organic chemistry class were only 25% accurate when asked to translate between these diagrammatic formats. In this research, it was also clear that students' main difficulty was with representing the 3-D structure of molecules. The most common error was a *spatial* error in which students drew a stereoisomer of the molecule to be drawn rather than the correct molecule. That is, students typically drew a molecule that was made up of the same atoms and had the same connectivity between atoms, but had a different 3-D arrangement of the atoms in space. In-depth interviews indicated that these students either failed to understand the importance of preserving the 3-D spatial arrangement of atoms in the translation (and hence the distinction between conformers and isomers) or lacked understanding of the diagram conventions (Padalkar & Hegarty, 2013). Moreover, students were overconfident in their ability to translate between diagrams. They believed that they were highly accurate, although they made spatial errors on most problems.

Previous research has also revealed that students can be more successful in translating between diagrams when they use models, but when provided with models, many students simply ignore them. In an initial study, Stull et al. (2012) made models available to students as they performed diagram translation tasks and found that most students ignored the models. In later studies, when they encouraged students to use the models, successful students aligned the model with the given diagram and then transformed it (by rotation, etc.) to align it to the diagram to be drawn, so that they used the model to externalize the relevant spatial transformations. However, many of the students still ignored the provided models. Students who used models had much better performance (ranging

[1] When the substituents on adjacent carbon atoms are at the maximum distance from each other, chemists say that the molecule is in a staggered conformation, and when the substituents on adjacent carbon atoms are in closest proximity, it is referred to as an *eclipsed conformation* (see Figure 2).

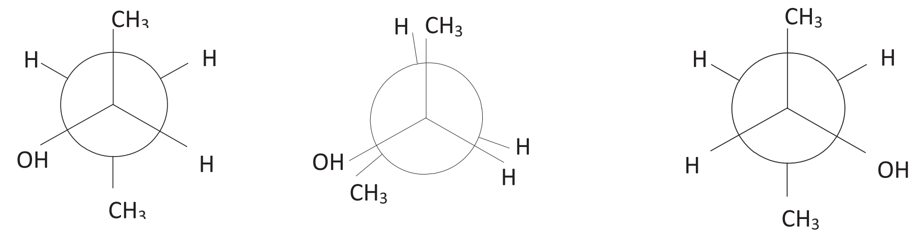


Figure 2a: (S)-2-butanol (staggered) **Figure 2b: (S)-2-butanol (eclipsed)** **Figure 2c: (R)-2-butanol (staggered)**

Figure 2. Examples of different conformations of the same molecule and an isomer of that molecule. Figure 2a and 2b are different (staggered vs. eclipsed) conformations of the same molecule. Figure 2c represents an isomer because the relative positions of the groups attached to the front carbon are different. Specifically, the positions of H and OH are switched.

from 45% to 66% accuracy in different experiments) compared with students in no-model control groups (around 25% accuracy). Students who had models available, but did not use them consistently, performed no better than controls. In later experiments, providing models that were already aligned with the given diagram or explicitly pointing out the correspondence between the molecular substituents in the given diagram and the model did not significantly increase use of the model or improve performance on the task. More generally, organic chemistry students have poor attitudes toward using models. Although students are encouraged to buy model kits and use them for homework, they are not required to use them, and a recent survey of chemistry students revealed that they rarely use models, even when encouraged to do so by their instructor (Steiff, Scopelitis, Lira, & DeSutter, 2013).

In summary, after taking an introductory course in organic chemistry, many students lack *representational competence* in that they are unable to preserve key spatial relations when translating between diagrams, and they lack *meta-representational competence* in that they do not use models, even though models can be helpful to them.

In designing an intervention to address these challenges, we considered why students might not use models, even though they are helpful. First, we argue that students need to use models and experience their benefits in order to discover how models can be helpful in tasks such as diagram translation. Although diagram translation becomes easier when external rotations of the model can replace internal (mental) rotations, students typically do not discover this model-based strategy on their own (Stull et al., 2012), possibly because the use of models to make spatial inferences is difficult. For example, to use a model, a student must understand the correspondence between the model and the entity it represents and that transformations of the model reveal something about its referent (Ainsworth, 2006). Research in other domains (and with children) has shown that a student who does not spontaneously discover or produce an effective strategy for a task can often be taught the strategy (e.g., Brown, Campione, & Day, 1981). Here, we propose that the same is true of adults who are novices in a domain. Moreover, adults spontaneously adopt better learning strategies, provided that they experience their benefits (E. L. Bjork, deWinstanley, & Storm, 2007; DeWinstanley & Bjork, 2004). Therefore, a critical aspect of our intervention was that students experienced the benefits of models.

Another possible reason for ignoring models is that students are overconfident in their ability to perform spatial inference tasks mentally (cf. Dunning et al., 2003), as documented by Padalkar and Hegarty (2013). Specifically, they might not realize that the 3-D locations of substituents are relevant to preserving the structure of the molecule in the diagram translation, so they believe they are accurate if they just maintain the connectivity. This is an example of an illusion of understanding or competence (R. A. Bjork, 1999; Rozenblit & Keil, 2002), which has been found to impede productive use of learning resources. Therefore, another desirable aspect of our intervention was that it should provide students with feedback on their performance, to address any illusions of understanding or competence.

Models as Feedback

In the intervention tested in Experiment 1, students first attempted to solve diagram translation problems and were then guided to use models to check their solutions so that the models provided feedback. The sequence of first producing solutions to problems, before being guided in the use of models, is consistent with the generation effect (Hirshman & Bjork, 1988; Slamecka & Graf, 1978) as a general learning principle, and more specifically with the idea that students may learn most effectively from models and other learning aids if they are active in learning and grapple with the problems in advance (Martin & Schwartz, 2005; Schwartz & Bransford, 1998; Schwartz & Martin, 2004). For example, in a recent article, Hinze et al. (2013) speculated that although students have difficulties adopting new representations in chemistry, this might allow them to identify gaps in their knowledge that prepare them for future learning. It is important to note that in our situation, generating a solution involved drawing diagrams. The benefits of drawing diagrams for science learning have been well documented (Ainsworth et al., 2011), and drawing diagrams is recommended in the National Science Education standards (National Research Council, 1996). A potential negative consequence of any generative activity is that students make errors, and based on previous research (Padalkar & Hegarty, 2013; Stull et al., 2012), we expected that students would initially draw incorrect diagrams. However, recent research has indicated that unsuccessful attempts on tests can be beneficial to learning, provided that they are followed by feedback (Kornell, 2014; Kornell, Hays, & Bjork, 2009).

A critical aspect of our intervention was that models were used to give students feedback on their solutions. Using models as feedback contrasts with most typical uses of models in instruction, in which models are used by the teacher to demonstrate concepts or by students to explore concepts or solve problems. *Feedback* can be defined as "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007, p. 81). The main features of feedback are that it occurs only after a student has attempted a task, and it provides an evaluation of performance on that task. Feedback is a powerful factor influencing students' achievement. For example, a meta-analysis of 196 studies revealed a mean effect size of 0.79 for feedback manipulations (Hattie & Timperley, 2007). At the same time, not all forms of feedback are equally effective (Shute, 2008). Feedback can be categorized as (a) feedback about a task or product (FT), that is, stating whether the work is correct or incorrect; (b) feedback about the processing of the task (FP), that is, providing processing information or alternative strategies; (c) feedback about self-regulation (FR), that is, providing information about self-regulatory proficiencies and self-beliefs; and (d) feedback about the self as a person (FS), which is unrelated to performance on a task. The first three kinds of feedback have been found to be effective in prior research, whereas the fourth is most common but least effective (Hattie & Timperley, 2007).

Our intervention in Experiment 1 incorporated the first two kinds of feedback. It provided students with feedback about the product of their task (FT) by informing them whether their diagram was correct and, if not, what mistake they had made. This type of feedback addresses possible illusions of competence. It also addressed the processing of the task (FP) by teaching students a strategy for using a model to check their solutions. This ensured that students experienced the benefits of models. A distinctive aspect of our feedback was that it emerged as a student worked with a model, so that the feedback was grounded in reality and discovered by the student, rather than being provided by a tutor. While a tutor was present to guide the student in the use of the models, the tutor did not tell the student whether his or her answer was correct or incorrect. The student discovered this in interaction with the model. We refer to this as *model-based feedback*.

Experiment 1

In Experiment 1, we tested our intervention in a study that used a pretest–posttest design with an experimental group who received the intervention and a control group who did not receive this intervention. On the basis of previous research (Stull et al., 2012), we expected that most students would not spontaneously adopt the model strategy and would have poor performance in the pretest. We predicted that participants who received the intervention would perform better than control participants in the posttest and that they would use the models more frequently in the posttest. We tested whether the improved performance was mediated by use of models. We also predicted that attitudes toward using models would improve for the intervention group after they had experienced the benefits of models but that attitudes would not change for controls.

Method

Participants and design. The participants were 54 undergraduate students at a research university who had completed at least one Organic Chemistry course. They received either course credit or $20 for their participation. The experiment followed a pretest–posttest design with control and experimental groups. The experimental group consisted of 30 participants (15 females), and the control group consisted of 24 participants (12 females). Six students (two in the experimental group and four in the control group) were not included in the analyses because they had perfect or almost perfect performance on the pretest (made either zero or one error), indicating that they had already mastered the task and there was nothing for them to learn. Students in the experimental and control groups did not differ in age, grade-point average, number of years in college, or number of organic chemistry courses completed (see Table 1).

Materials. In the main experimental task, participants were given one kind of diagram of a molecule (Dash-wedge, Newman

Table 1
Demographics of Students in the Different Conditions of Experiments 1 and 2

Variable	Experimental *M* (*SD*)	Control *M* (*SD*)
Experiment 1		
Age	20.33 (.96)	20.17 (1.61)
GPA	2.88 (.86)	3.09 (.75)
Years in college	3.10 (.55)	2.96 (.62)
No. of Organic Chemistry classes	2.43 (.77)	2.04 (.96)

	Match-models *M* (*SD*)	Feedback *M* (*SD*)	Control *M* (*SD*)
Experiment 2			
Age	20.58 (.95)	20.32 (2.16)	20.09 (.97)
GPA	2.94 (.47)	3.13 (.35)	3.09 (.41)
Years in college	2.67 (.76)	2.86 (.77)	2.82 (.80)
No. of Organic Chemistry classes	2.13 (.83)	2.27 (.83)	2.09 (.92)

Note. GPA = grade-point average; No. = Number.

projection, or Fischer projection) and were asked to draw one of the other two kinds of diagram of the same molecule. Thus, there were six kinds of problems (translate Dash-wedge to Newman, Newman to Fischer, Fischer to Dash-wedge, and vice versa). The pretest included six problems (one of each kind) with four-carbon molecules. The posttest included a second set of six problems with four-carbon molecules (mirror images of the molecules in the pretest problems) and six additional problems with five-carbon molecules. All of the molecules used in the problems had two chiral carbons that were linked by a molecular bond. The worksheets for this representation translation task (8.5″ × 11″ sheets of paper) included an instruction on the top (stating which diagram to draw) and a diagram below it. The solution space on the worksheet for the pretest was divided into two equal parts by a horizontal line, and participants were asked to draw their solution above the line. Posttest worksheets were not divided, and participants were allowed to draw their solution anywhere below the given diagram.

Concrete (ball-and-stick) models were provided for each problem. The models were constructed from a commercial molecular modeling kit (HGS Introductory Organic Chemistry Set 1000) that is commonly used in high school and college chemistry courses. The models were presented in the same conformation as the given diagram, but they were not aligned with it.

The pretest was followed by a short questionnaire on participants' level of confidence in their solutions and the usefulness of the concrete models. It contained six statements (see Table 2). Participants marked their agreement with each on a scale ranging from 1 (*strongly agree*) to 5 (*strongly disagree*). The posttest was followed by a questionnaire that included demographics questions and the same statements about confidence and usefulness of models as the pretest questionnaire.

Procedure. Both experimental and control groups were first given basic instructions, which included the nature of the task, examples of the three kinds of diagrams, and reminders of the conventions of each diagram (see the Appendix). After they read the instruction sheet, they were told that it would be kept face down on the table and they could refer to it as necessary. They were also given a concrete (ball-and-stick) model and reminded of the color codes for the different atoms in the model. The model was positioned in a clay stand, and the experimenter demonstrated that the concrete model could be taken out of the stand and that it could be rotated in space and around the main carbon-carbon bond, linking the two chiral carbons of the molecules. They were told that they could draw any conformation of the molecule.

Both groups first solved the pretest problems and responded to the short questionnaire on confidence in their solutions and the usefulness of the models. Then the experimental group went through a training intervention (described below), and the control group participants were given a 5-min break. Next, both groups solved the posttest items and responded to the posttest questionnaire (including the demographics reported in Table 1, confidence and usefulness of models). Participants were videotaped with their consent during the drawing task

Intervention. The intervention included directions for participants to use models to check their own solutions to the six pretest problems and to draw correct solutions if any of their solutions were found to be incorrect. First, participants reviewed their solution to the pretest problems, one at a time. For each problem, they were provided with the concrete model of the molecule in the problem, and their solution was checked in three steps. The first step was to match the model with the given diagram (i.e., put the model in the same orientation and conformation as the given diagram). This gave participants an opportunity to confirm that the model indeed represented the same molecule as the given diagram and also gave practice in seeing the correspondence between the model and diagram. In the second step, participants were asked to align the model with their solution to the problem. This was possible only if the participant had drawn a correct solution. Once the participant aligned the model with the solution, he or she was asked to move to the next problem. (If they did not align it correctly, the experimenter drew their attention to substituents that did not match.) If the solution was incorrect, it was not possible to structurally align the model with the solution, and the participant discovered that his or her solution was incorrect. The third step involved drawing a new corrected solution (below the horizontal line). If the participant again drew an incorrect solution, Steps 2 and 3 were repeated.

Table 2
Participants' Self-Reports About Their Confidence and Perception About Models

Item	Condition	Pretest *M* (*SD*)	Posttest *M* (*SD*)	*p*
1. I am confident about my solutions.	Experimental*	2.57 (0.92)	1.68 (0.67)	<.001
	Control	2.55 (0.76)	2.63 (0.78)	.69
2. The transformation problems were challenging.	Experimental	3.04 (0.84)	2.79 (0.88)	.15
	Control	2.70 (0.73)	2.80 (0.95)	.73
3. The molecular models were helpful.	Experimental*	2.68 (1.36)	1.14 (0.37)	<.001
	Control	2.20 (1.36)	1.90 (1.29)	.45
4. I did not need to use the models.	Experimental*	2.93 (1.30)	4.43 (0.79)	<.001
	Control	3.60 (1.27)	3.75 (1.25)	.73
5. The models helped me visualize the projections.	Experimental*	2.59 (1.40)	1.36 (0.68)	<.001
	Control	2.05 (1.05)	1.65 (0.93)	.34
6. I found it necessary to pick up the models during the task.	Experimental*	3.68 (1.22)	1.68 (1.34)	<.001
	Control	2.50 (1.57)	2.20 (1.47)	1.00

Note. Ratings were made on a 1–5 scale, where 1 = *Strongly agree* and 5 = *Strongly disagree*. An asterisk indicates a significant difference from pre- to posttest for an item and group; *p* values are exact significance based on a sign test.

Out of 180 solutions, 47 were correct in the pretest, 114 were corrected in the first cycle of the intervention, and 17 were corrected in the second cycle of the intervention. Finally, there were two cases in which participants drew an incorrect solution on the third attempt. In both cases, the participant was told his or her mistake and was asked to go to the next problem.

Coding. The diagram translation problems were coded in two ways. First, each item was scored as correct (1) or incorrect (0). To receive a score of 1, a drawing had to show one of the conformations of the molecule using the conventions of the diagram to be drawn. Interrater reliability for this coding was 98.9%. The proportion of correct solutions served as the accuracy score for the pretest and posttest problems.

Second, errors were classified as (a) spatial errors, (b) connectivity errors, or (c) fundamental errors. In the case of spatial errors, the drawn diagrams were made up of the correct molecular substituents and the connectivity between these substituents was correct, but their 3-D spatial arrangement was incorrect (in some cases, the student drew a diagram that depicted an arrangement of the substituents that was physically impossible).[2] Connectivity errors were solutions in which the diagram drawn was made up of the correct substituents, but these were connected to the wrong chiral carbon atoms. Finally, fundamental errors included drawing the wrong type of diagram, drawing the diagram incorrectly, or drawing a diagram with missing or extra substituents. Sample errors are shown in Figure 3. Data for 20 participants (37% of the data) were coded independently by two researchers to establish interrater reliability. The agreement between the two raters for type of error was 93.9%. Discrepancies were resolved by consensus.

Coding of model use. Model use was coded from the videos by two experienced coders for the following behaviors.

1. Aligning the model to the orientation of the starting diagram (align-start).

2. Aligning the model with the orientation and conformation of the target diagram that the participant drew (align-target).

The measure of model use was the proportion of trials on which the behavior occurred. Interrater reliability for the two coders was 89.9%, with discrepancies resolved by consensus.

Results

Accuracy. Performance of the two groups on the pretest and posttest problems is shown in Figure 4. Accuracy of solutions for the four-carbon problems (mean proportion correct = .65, SD = 0.33) and five-carbon problems (M = 0.68, SD = 0.32) in the posttest did not differ significantly, $t(47) = 1.278$, $p = .21$, so posttest accuracy is calculated for all 12 problems together.

We predicted that student performance would increase from pretest to posttest and that the intervention group would outperform the control group at posttest. A 2 (time of testing: pretest, posttest) × 2 (condition: intervention, control) analysis of variance revealed significant main effects of time of testing, $F(1, 46) = 96.04$, $p < .001$, $\eta_p^2 = .68$, and condition, $F(1, 46) = 5.48$, $p = .02$, $\eta_p^2 = .11$, and an interaction of time and condition, $F(1, 46) = 36.23$, $p < .001$, $\eta_p^2 = .44$. As Figure 4 shows, performance was relatively poor at pretest, and simple effects analyses revealed no significant difference between the intervention and control conditions, $F(1, 46) = 2.69$, $p > .10$, as expected. In contrast, the intervention group was significantly more accurate than the control group at posttest, $F(1, 46) = 23.57$, $p < .001$, $\eta_p^2 = .34$, as predicted. Specifically, on the posttest, the intervention group drew correct diagrams on 9.8 (SD = 2.78) of the 12 trials, compared with 5.4 (SD = 3.5) for the control group. Although both groups significantly improved from pretest to posttest, the effect size was much greater for the intervention group (d = 2.08), $t(27) = 10.92$, $p < .001$, than for the control group (d = .77), $t(19) = 3.12$, $p = .006$.

Error analysis. The percentages of trials that were classified as spatial errors, connectivity errors, and fundamental errors are shown in Table 3. Consistent with previous research (Padalkar & Hegarty, 2013; Stull et al., 2012), the majority of errors in the pretest were spatial errors indicating poor understanding of the 3-D configuration of components. The pattern of spatial errors mirrored the results for overall errors. Spatial errors decreased significantly from the pretest to the posttest, $F(1, 46) = 75.36$, $p < .001$, $\eta_p^2 = .62$; participants in the intervention condition made fewer spatial errors than the control group, $F(1, 46) = 10.66$, $p = .002$, $\eta_p^2 = .19$; and there was a significant interaction of time (pretest, posttest) with intervention, $F(1, 46) = 21.47$, $p < .001$, $\eta_p^2 = .32$, such that spatial errors decreased more for the intervention group than the control group. In contrast, connectivity and fundamental errors were rare (see Table 3), indicating that students had a good understanding of the diagram formalisms and how they show connectivity. Fundamental errors decreased significantly from the pretest to the posttest, $F(1, 46) = 9.03$, $p = .004$, $\eta_p^2 = .15$, but did not differ for the intervention groups ($p > .20$ for both main effect and interaction). There were no significant effects of either time (pretest, posttest) or condition (intervention, control) on connectivity errors.

Model use. The strategy trained in the intervention was to first align the model with the given diagram (to confirm that they represented the same molecule), then manipulate it to the orientation and conformation of the diagram to be drawn, and finally draw the required diagram. As shown in Figure 5a, aligning the model with the given diagram was uncommon in the pretest, indicating that, as expected, students did not spontaneously use the model. Participants in the intervention group aligned the model with the given diagram more often in the posttest than the pretest, $t(27) = 3.8$, $p = .001$, whereas the control group did not differ in this behavior from pretest to posttest, $t(19) = 1.69$, $p = .13$.

Similarly, the percentage of trials on which participants aligned the model with the perspective and conformation of the diagram that they eventually drew is shown in Figure 5b. This behavior indicates that students were performing the relevant rotations externally on the model. Again, this behavior was uncommon in the pretest. Although aligning the model to the target diagram was more common in the posttest than the pretest among both groups, the effect size for the difference from pretest to posttest was much larger for the experimental group (Cohen's d = 3.057), $t(27) = 10.26$, $p < .001$, than for the control group (d = .883), $t(19) = 3.34$, $p = .003$.

Accuracy was highly correlated with aligning the model with the target diagram at both pretest (r = .70) and posttest (r = .83). To examine whether model use mediated the significant

[2] When attempting to draw a Dash-wedge diagram, they drew a solid line in between the dash and the wedge (see Figure 3).

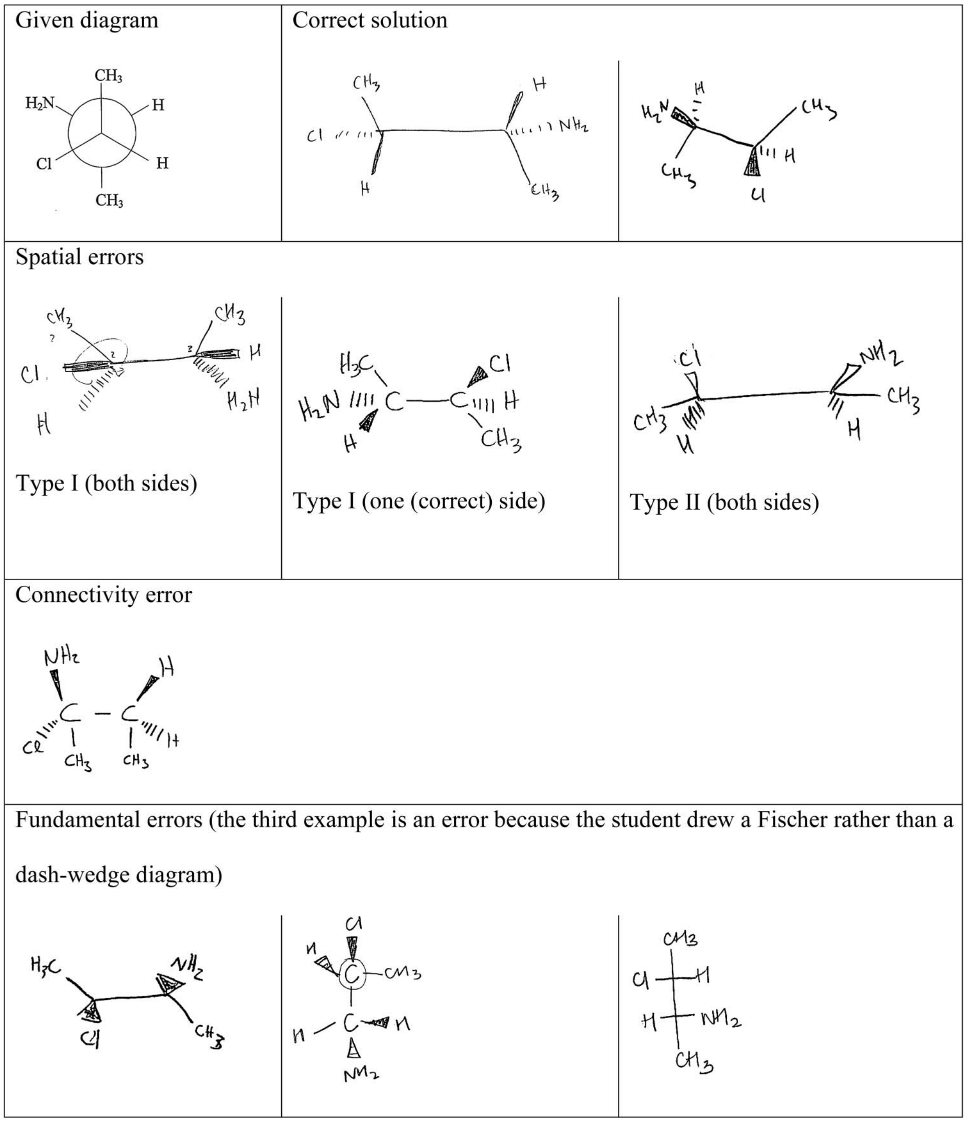


Figure 3. Examples of correct solution and incorrect solutions for a trial in which students were given a Newman projection and asked to draw a Dash-wedge diagram of the same molecule. In the case of fundamental errors (the third example is an error because the student drew a Fischer rather than a Dash-wedge diagram).

difference in accuracy between the intervention and control groups on the posttest, we conducted a mediation analysis (Baron & Kenny, 1986). The results of this analysis are shown in Figure 6. First, accuracy was regressed on experimental condition (intervention, control) and revealed a significant effect ($\beta = .58, p < .001$). Second, model use was regressed on experimental condition and again revealed a significant effect, consistent with our previous results ($\beta = .59, p < .001$). The third step was to examine the effect of the mediator (model use) on the dependent measure (accuracy), and this effect was also significant ($\beta = .83, p < .001$). Finally, we assessed the effect of the independent variable (intervention vs. control) on the dependent measure (accuracy) while controlling for the measure of model use. In this analysis, the path from the independent measure and dependent measure was not significant ($\beta = .14, p = .17$), indicating that model use mediated the relationship between condition (intervention vs. control) and accuracy on the posttest.

Attitudes toward models and confidence level. Table 2 shows participants' average confidence level and attitudes toward models based on the questionnaire data. The significance level (p value) was based on a sign test (a nonparametric test was used as the data were interval scale). Students' confidence level (Statement 1) and perceived usefulness of models (Statements 3, 4, 5, and 6) significantly increased for the experimental participants after the intervention. Interestingly, there was a trend for these students to rate the problems to be more challenging after the intervention than before, although their performance and confidence level was higher. This suggests that they gained an understanding of the complexity of the problems after they understood how to solve

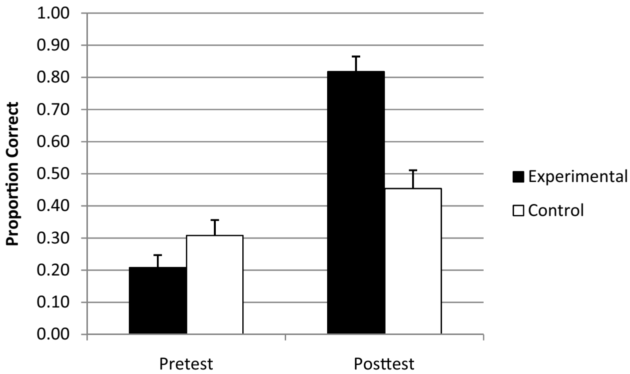
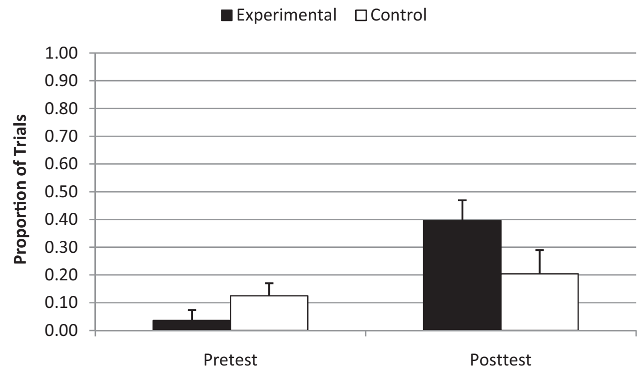


Figure 4. Average proportion of correct solutions before (pretest) and after the intervention (posttest) in Experiment 1. Error bars show standard error of the mean.

(a) Align Start



(b) Align Target

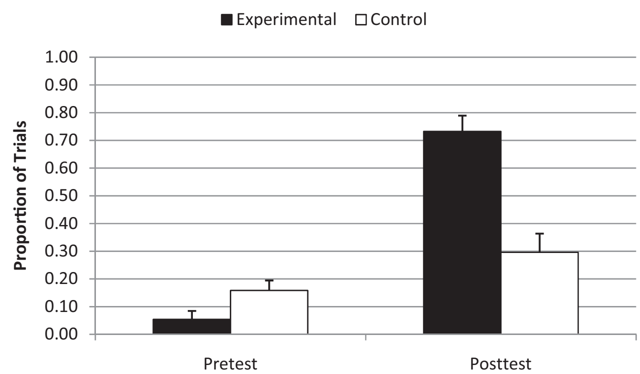


Figure 5. Proportion of trials on which participants (a) aligned the model with the given diagram (align-start) and (b) aligned the model with the target diagram that they drew (align-target) before and after the intervention. Error bars show standard error of the mean.

them. There were no significant differences between pretest and posttest ratings for any of the statements among the control group.

Discussion

As predicted, the experimental group performed significantly better than the control group on the posttest. The participants in the experimental group used the models more frequently and meaningfully and judged that models were more useful after the intervention. Increases in performance from the pretest to the posttest were mediated by use of the models and were accompanied by improvements in attitudes toward models. In conclusion, the intervention was effective in teaching students to perform the diagram translation task successfully using concrete models.

It is perhaps not surprising that the intervention in Experiment 1 was successful, as it combined two learning principles. First, it provided feedback to the participants, which addressed any overconfidence or illusions of understanding or any failure to understand that it was important to preserve the 3-D spatial relations. Second, it made students enact the spatial transformation with a model, revealing how the model could be used to help translate between the diagrams, so that they experienced the benefits of models. If participants are overconfident or do not understand the importance of preserving the 3-D spatial relations, and those are the only factors limiting performance, then giving them constructive feedback that they drew an isomer, rather than the correct molecule, should be sufficient to improve performance. In contrast, if they are unable to discover the model-based strategy on their own and just need to have experience using the models, then giving them experience in matching concrete models to diagrams and manipulating models should improve their performance. To identify which aspects of the intervention were most effective, we conducted Experiment 2 in which we compared a feedback-only condition and a "match-model" condition in which students did not receive explicit feedback.

Experiment 2

Aspects of the intervention in Experiment 1 were separated into two different interventions in Experiment 2. There were three

Table 3
Mean Percentage of Incorrect Trials on Which Participants Made a Fundamental Error, Connectivity Error, and a Spatial Error in Experiment 1

	Pretest			Posttest		
Variable	Fundamental	Connectivity	Spatial	Fundamental	Connectivity	Spatial
Experimental	13.69 (17.60)	2.38 (9.85)	63.69 (21.78)	4.17 (7.35)	2.38 (4.45)	11.61 (16.09)
Control	7.50 (13.76)	1.67 (5.13)	60.00 (19.79)	3.33 (8.29)	7.08 (12.17)	44.17 (23.12)

Note. Standard deviations appear in parentheses.

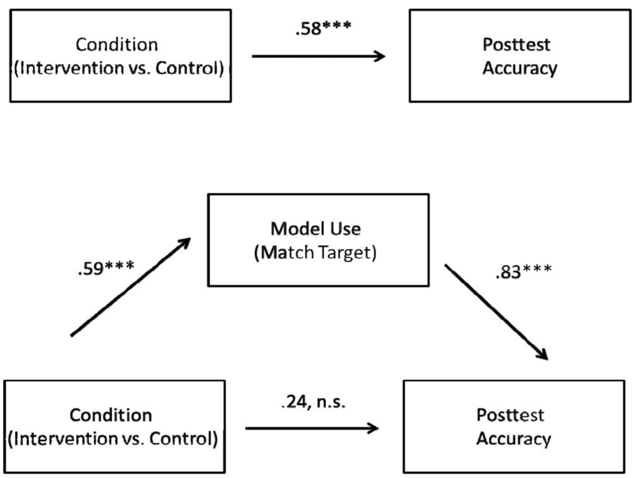


Figure 6. Results of the mediation analysis for Experiment 1. *** $p < .001$.

conditions, namely, a control condition, a feedback condition in which students received verbal feedback on their solutions but did not align models to diagrams, and a match-model condition in which students practiced aligning models to diagrams but did not receive feedback on their pretest drawings. The feedback condition provided the students with an evaluation of their pretest performance. This feedback was on the product of the task (FT; cf. Hattie & Timperley 2007) but did not teach them a process for performing the task (FP). In contrast the match-model gave students practice in matching diagrams to models, but not in the context of any evaluation of their pretest performance. Although this group did not get feedback on their pretest performance, they received instruction on how to align the models, which addressed errors they had made in the pretest condition.

If participants have illusions of understanding (cf. R. A. Bjork, 1999), that is, are overconfident, or do not understand the importance of preserving the 3-D spatial relations in the transformation (see Figure 2), then the feedback should make it clear to them that they did not understand the task and that the 3-D spatial relations are relevant. However, this might or might not be sufficient to improve performance relevant to a control group, depending on whether they can develop an effective strategy for performing the task. If participants lack the spatial skills required to discover the model-based strategy on their own, it is possible that having them match models to diagrams will enable them to experience the benefits of models. This experience might or might not improve use of models in the diagram translation task, depending on whether they see the relevance of the model-matching activity to performing this task. Because the match-model condition explicitly instructed students to align model with the diagrams, we expected that use of models would increase and attitudes toward models would improve in this condition. As in Experiment 1, we tested the hypothesis that significant differences between the groups in posttest performance would be mediated by the use of models.

Method

Participants and design. The participants were 75 undergraduate students recruited from a research university. All participants had completed at least one course in organic chemistry. The experiment followed a 2 (pretest posttest) $\times$ 3 (model, feedback, control) design. The participants were randomly assigned to one of the three groups with the constraint that the proportions of males and females remained approximately equal in the three groups. There were 25 students in each group, with 13 females in the models group, 13 females in the feedback group, and 14 females in the control group. A small number of participants (one in the models group, three in the feedback group, and three in the control group) were not included in the analyses because they made either zero or one error on the pretest, indicating that they had already mastered the task and there was nothing for them to learn. Students in the experimental and control groups did not differ significantly in age, grade-point average, number of years in college, or number of organic chemistry courses completed (see Table 1). Participants received either course credit or $20 for their participation.

Materials. The problems used in this experiment were the same as in Experiment 1. The only difference was that the pretest problem sheets were not divided by a horizontal line. The demographic questionnaire was administered on a computer, in contrast with the previous experiment, in which it was administered on paper.

Procedure. As in Experiment 1, participants in all three groups were reminded of the conventions of the three diagrammatic representations and the concrete models (which they had learned in their chemistry classes; see the Appendix) and were shown how they could manipulate the model. Then they solved the same six pretest problems as in Experiment 1, with models present. After this, the participants answered the demographic questionnaire and while they were doing this, the experimenter checked the solutions of the pretest problems. Next, the feedback group was given feedback on their pretest problems, and the match-model group received the model intervention, as described below. Then all participants solved the same 12 posttest problems as in Experiment 1 (the control group solved the posttest problems immediately after they completed the questionnaire). Finally, participants completed the short questionnaire about confidence and attitudes toward models.

Interventions. Participants in the feedback group were given information on the accuracy of their solutions and (if relevant) the mistakes that they made. An example of feedback for a spatial error was "You have switched positions of these two items (substituents), so you have drawn an isomer of the given diagram." For a connectivity error, the participants were told "In the given diagram, these three groups are attached to one carbon and these three are attached to another. In your diagram this group should have been here and this group should have been here" (the experimenter pointed to the correct locations as she said "here"). An example of feedback for a "fundamental error" is "Notice that there is only one . . . group in the given diagram. You have drawn two." The corresponding model was kept available while the feedback was provided.

Participants in the match-model group were not provided any evaluation of their pretest problems. Instead, they were given six printed diagrams (two of each kind, three on each page). An unaligned model was provided with each diagram, and participants were asked to attempt to align the model with the diagram and to judge whether the diagram and the model represented the same organic molecule or different organic molecules. If they did not align the model correctly, the experimenter helped them by pointing to the correct conventions in the instruction sheet. Three of the models matched and three were isomers of the depicted diagram (containing the same molecular substituents but in a different spatial arrangement). In four of the six cases, participants had to rotate the model around the main carbon-carbon bond (i.e., reconfigure from Staggered to Eclipsed, or vice versa) in order to match the model with the diagram. Note that each match-model trial involved comparing two representations (a model and a diagram), neither of which was created by the participants.

Results

Accuracy. Performance of the three groups on the pretest and posttest problems is shown in Figure 7. A 2 (time of testing: pretest, posttest) × 3 (condition: match-model, feedback, control) analysis of variance revealed significant main effects of time of testing, $F(1, 65) = 36.45$, $p < .001$, $\eta_p^2 = .36$; condition, $F(2, 65) = 3.19$, $p = .05$, $\eta_p^2 = .09$; and the interaction of time and condition, $F(2, 65) = 3.62$, $p = .03$, $\eta_p^2 = .10$. Simple effects analyses revealed no significant difference between the three conditions at pretest, $F(2, 65) = 1.22$, $p > .30$. In contrast, there was a significant difference between the groups at posttest, $F(2, 65) = 4.43$, $p = .02$, $\eta_p^2 = .12$. Pairwise comparisons (with Bonferroni correction) indicated that the match-model group was more accurate than the control group at posttest ($p = .01$), whereas the feedback group did not differ significantly from either the control or match-model groups ($p > .16$ in both cases). Although all three groups improved from pretest to posttest, the effect size was greater for the match-model group (Cohen's $d = 1.12$), $t(23) = 5.35$, $p < .001$, than the feedback group (Cohen's $d = .70$), $t(21) = 3.28$, $p = .004$, and the control group (Cohen's $d = .56$), $t(21) = 2.60$, $p = .017$.

Error analysis. Incorrect solutions were coded as fundamental errors, connectivity errors, and spatial errors as in Experiment 1 (see Table 4). Again, most errors in the pretest were spatial errors, indicating poor understanding of the 3-D configuration of components. Spatial errors decreased significantly from the pretest to the posttest, $F(2, 65) = 30.35$, $p < .001$, $\eta_p^2 = .32$. The main effect of experimental condition was not significant, $F(2, 65) = 1.92$, $p = .15$, but the critical interaction of Time of Testing (pretest, posttest) × Condition (feedback, model, control) was observed, $F(2, 65) = 4.17$, $p = .02$, $\eta_p^2 = .11$. As Table 4 shows, spatial errors decreased most for the match-model group and least for the control group. Connectivity and fundamental errors were rare (see Table 4), indicating that students had a good understanding of the diagram formalisms and how they show connectivity. Fundamental errors decreased significantly from the pretest to the posttest, $F(2, 65) = 9.76$, $p = .003$, $\eta_p^2 = .13$, but did not differ for the intervention groups ($F > 1.0$ for both main effect and interaction). There were no significant effects of either time (pretest, posttest) or condition (feedback, match-model, control) on connectivity errors.

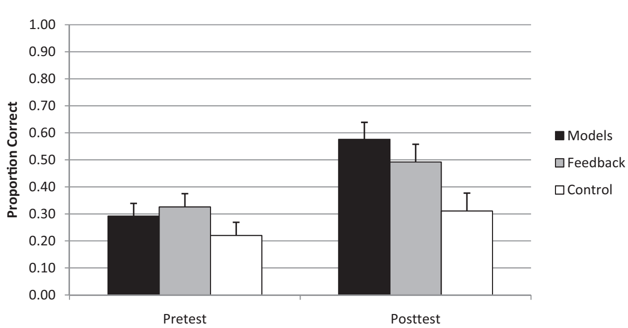


Figure 7. Average proportion of correct solutions before (pretest) and after the interventions (posttest) in Experiment 2. Error bars show standard error of the mean.

Model use. The percentage of trials on which the model was aligned to the given diagram (align-start) is shown in Figure 8a. A 2 (time of testing: pretest, posttest) × 3 (condition: feedback, match-model, control) analysis of variance revealed significant main effects of time of testing, $F(1, 65) = 71.85$, $p < .001$, $\eta_p^2 = .53$; condition, $F(2, 65) = 5.01$, $p < .01$, $\eta_p^2 = .13$; and the interaction of time and condition, $F(2, 65) = 17.15$, $p < .001$, $\eta_p^2 = .35$. Simple effects analyses revealed no significant difference between the three conditions at pretest, $F(2, 65) < 1$. In contrast, there was a significant difference between the groups at posttest, $F(2, 65) = 9.09$, $p < .001$, $\eta_p^2 = .22$. Pairwise comparisons (with Bonferroni correction) indicated that at posttest, the match-model group was more likely to match the model with the given diagram than both the feedback group ($p = .02$) and the control group ($p < .001$), whereas the latter two groups did not significantly differ in this behavior.

The same pattern emerged when we examined whether students aligned the model to match the perspective and conformation of the diagram to be drawn (see Figure 8b). A 2 (time of testing: pretest, posttest) × 3 (condition: feedback, match-model, control) analysis of variance revealed significant main effects of time, $F(1, 65) = 58.15$, $p < .001$, $\eta_p^2 = .47$; condition, $F(2, 65) = 7.59$, $p = .001$, $\eta_p^2 = .19$; and the interaction of time and condition, $F(2, 65) = 7.60$, $p < .001$, $\eta_p^2 = .19$. Simple effects analyses revealed no significant difference between the three conditions at pretest, $F(2, 65) = 1.63$, $p = .20$. In contrast, there was a significant difference between the groups at posttest, $F(2, 65) = 6.41$, $p = .003$, $\eta_p^2 = .17$. Pairwise comparisons (with Bonferroni correction) indicated that at posttest, the match-model group was more likely to match the model with the diagram to be drawn than the control group ($p = .003$), and the feedback group did not significantly differ from either the match-model or control group.

Aligning the model to match the target diagram was highly correlated with accuracy at both pretest ($r = .70$) and posttest ($r = .84$). To examine whether model use mediated the significant difference in accuracy between the match-model and control groups on the posttest, we conducted a mediation analysis (Baron & Kenny, 1986). The results of this analysis are shown in Figure 9. First, accuracy was regressed on experimental condition (match-model, control) and revealed a significant effect ($\beta = .44$, $p = .002$). Second, model use was regressed on experimental condition and again revealed a significant effect, consistent with our previous results ($\beta = .46$, $p = .001$). The third step was to examine the effect of the mediator (model use) on the dependent measure (accuracy), and this effect was also significant ($\beta = .86$, $p < .001$). Finally, we assessed the effect of the independent variable (intervention vs. control) on the dependent measure (accuracy) while

Table 4
Mean Percentage of Incorrect Trials on Which Participants Made a Fundamental Error, Connectivity Error, and a Spatial Error in Experiment 2

	Pretest			Posttest		
Variable	Fundamental	Connectivity	Spatial	Fundamental	Connectivity	Spatial
Control	9.85 (13.27)	1.52 (4.90)	66.67 (20.57)	3.79 (8.02)	6.82 (10.17)	59.09 (23.42)
Feedback	5.30 (11.94)	1.52 (4.90)	60.61 (24.96)	2.27 (4.59)	2.27 (4.59)	49.24 (27.93)
Match-model	6.25 (10.78)	2.08 (7.47)	62.50 (22.12)	3.13 (5.39)	1.39 (4.01)	37.15 (28.87)

Note. Standard deviations appear in parentheses.

controlling for the measure of model use. In this analysis, the path from the independent measure to the dependent measure was not significant ($\beta = .06$, $p = .51$), indicating that model use mediated the relationship between condition (match-model, control) and accuracy on the posttest.

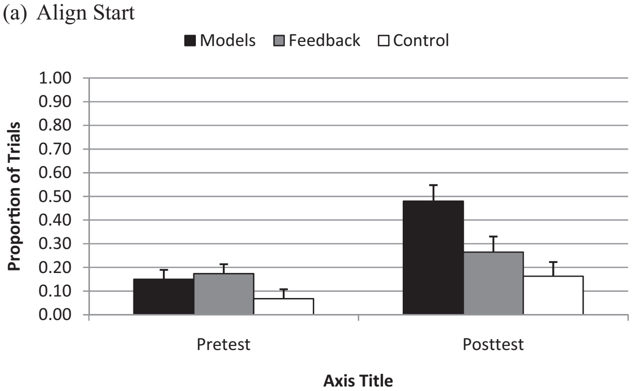
Attitudes toward models and confidence level. Table 5 shows shifts in attitudes from the pretest to the posttest based on the questionnaire data. On the basis of a sign test, confidence (Statement 1) significantly increased for the match-model group after the intervention. The perceived usefulness of models (Statements 4, 5, and 6) increased in both the feedback and the match-model groups, although the difference for Statement 5 was only marginally significant in the feedback condition. Explicit statements about the perceived difficulty of the problems and helpfulness of models (Statements 2 and 3) did not change significantly for the model and feedback groups from before to after the intervention. There were no significant differences between pretest and posttest ratings for any of the statements among the control group.

Discussion

The results of Experiment 2 indicated that the match-model group improved the most, and only this group significantly outperformed the control group on the posttest. These results suggest that failure on this task is not merely due to overconfidence or lack of understanding that the 3-D structure is relevant. If this were true, then the verbal feedback should have been sufficient to improve performance. Rather, the results suggest that students need to experience the benefits of models, as they did in the match-model condition, in order to discover the model-based strategy for diagram translation.

The feedback group did not use models significantly more than the control group, suggesting that this group could not use the concrete models effectively, even after they realized that they had made spatial errors. Interestingly, the questionnaire data indicated that this group rated the perceived usefulness of models as higher in the posttest, although they did not use models more than

(a) Align Start



(b) Align Target

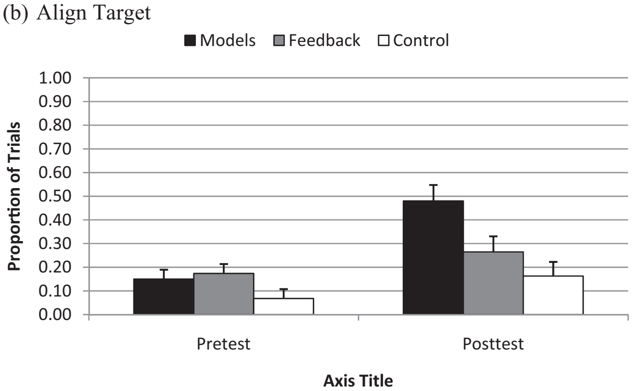


Figure 8. Proportion of trials on which the model was a: aligned to the given diagram (align-start) and b: aligned to the solution (align-target) before and after the interventions in Experiment 2. Error bars show standard error of the mean.

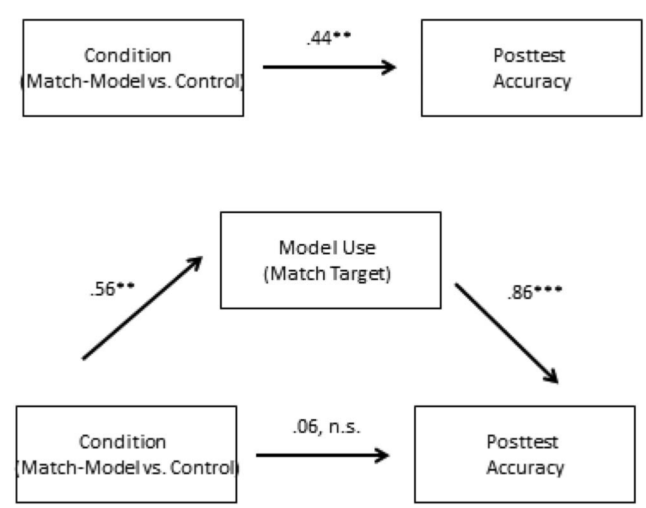


Figure 9. Results of the mediation analysis for Experiment 2. ** $p < .01$. *** $p < .001$.

Table 5
Participants' Self-Reports About Their Confidence and Perception About Models

Item	Condition	Pretest M (SD)	Posttest M (SD)	p
1. I am confident about my solutions.	Match-models*	2.79 (1.02)	2.33 (.92)	<.05
	Feedback	2.41 (.85)	2.32 (.89)	.80
	Control	2.23 (.81)	2.32 (.65)	.38
2. The transformation problems were challenging.	Match-models	2.83 (.87)	2.62 (.71)	.63
	Feedback	2.82 (.96)	2.64 (.85)	.55
	Control	2.77 (1.02)	2.64 (.95)	.22
3. The molecular models were helpful.	Match-models	2.00 (.98)	1.62 (.71)	.18
	Feedback	2.45 (1.14)	2.09 (.81)	.27
	Control	2.50 (.97)	2.27 (.16)	.39
4. I did not need to use the models.	Match-models*	3.33 (.96)	3.92 (.93)	.02
	Feedback*	3.09 (1.19)	3.86 (.99)	<.01
	Control	2.95 (1.05)	3.18 (1.26)	.55
5. The models helped me visualize the projections.	Match-models*	1.92 (.83)	1.46 (.58)	.02
	Feedback	2.50 (1.26)	1.95 (.99)	.07
	Control	2.45 (1.26)	2.41 (1.22)	1.00
6. I found it necessary to pick up the models during the task.	Match-models*	2.46 (1.02)	1.79 (.83)	<.01
	Feedback*	3.18 (1.43)	2.36 (1.18)	<.001
	Control	2.95 (1.43)	2.95 (1.36)	1.00

Note. Ratings were made on a 1–5 scale, where 1 = *Strongly agree* and 5 = *Strongly disagree*. An asterisk indicates a significant difference from pre- to posttest for an item and group; p values are exact significance based on a sign test.

controls. This suggests that being made aware of their errors led them to believe that models could help, but they could not spontaneously develop the model-based strategy. In contrast, although participants in the match-model group were not given any feedback on the accuracy of their pretest drawings, exposure to mapping the models and diagrams led them to develop a strategy for solving the diagram translation task using concrete models. Thus, experiencing the benefits of models was sufficient to increase use of models on the diagram translation task and to improve performance on the posttest, relative to a control group. These effects were accompanied by conscious judgments about the usefulness of models (as revealed by the questionnaire data). Thus, experiencing the benefits of models seems to be a better remedy than providing students feedback on the nature of their errors.

General Discussion

The aim of these experiments was to design and test interventions to improve aspects of students' representational and meta-representational competence in the domain of organic chemistry. Specifically, they addressed an aspect of representational competence, namely, that after taking an introductory course in organic chemistry, many students are unable to preserve key spatial relations when translating between diagrams (Keig & Rubba, 1993; Kozma & Russell, 2005). They also addressed aspects of meta-representational competence, namely, that students have poor attitudes toward models and when they are provided with models, they do not use them, even though models can be helpful to them (Stull et al., 2012).

In the intervention in Experiment 1, students' checked their solutions to the pretest problems using concrete models. This gave them feedback about their performance and presented an opportunity to map concrete models and diagrams. The intervention was effective in that the intervention group had more accurate performance on the posttest compared with a control group. In the pretest, most errors were *spatial* errors, due to lack of understanding of the 3-D structure of an organic molecule. These errors were significantly reduced in the posttest, indicating that we had improved students' representational competence. After the intervention, the students used the models more frequently and appropriately than the control group, and this use of the model mediated the effects of the intervention on posttest performance. Finally, students' attitudes toward models improved. Both their increased use of models and their improved attitudes indicate changes in meta-representational competence.

One possible interpretation of Experiment 1 is that it merely showed that if you teach students a strategy for performing a task, they will learn this strategy. However, the intervention in Experiment 1 did not explicitly teach students a strategy. Rather, it required students to match models to their pretest drawings so that they received feedback on their pretest drawings and experienced the benefits of models, and as a result they adopted the strategy of using models to perform the task. To test the effectiveness of instruction in matching models versus feedback, we separated these factors in a second experiment. In Experiment 2, the feedback group received an evaluation of their pretest drawings (without getting any opportunity to map concrete models and diagrams). In contrast, the match-model group aligned concrete models to diagrams to evaluate their equivalence (without receiving any feedback on their pretest performance, although requiring them to match models to diagrams may have suggested a strategy for performing the task). The match-model condition (Cohen's d = 1.12) was more effective in improving performance than the feedback condition (d = .70). Moreover, on the posttest, the feedback group was not significantly more accurate than a control group, who received no intervention. However, neither of these was as effective as intervention in the original intervention in Experiment 1 (d = 2.08). These results suggest that providing feedback without an opportunity to experience the benefits of models and providing an opportunity to develop the model strategy (a teaching technique) without providing feedback have limited success as compared with combination of these two. Although we should be cautious in comparing across studies, we note that participants in the two studies came from the same population and were comparable in terms of prior knowledge, academic achievement, and pretest performance (see Table 1 and Figures 4 and 7). However, students who received the intervention in Experiment 1 were over 80% correct on the posttest, whereas students in the match-model group of Experiment 2 were less than 60% correct, and those in the feedback condition of Experiment 2 were less than 50% accurate on the posttest.

The intervention in Experiment 1 combined several cognitive learning principles, including the generation effect, addressing illusions of understanding, and feedback. These principles have been tested primarily in the context of verbal learning and fact learning (R. A. Bjork, 1999; Dunning et al., 2003; Hirshman & Bjork, 1988; Kornell et al., 2009; Slamecka & Graf, 1978), so one contribution of this research is that it extends research on these basic principles to a complex spatial domain (see Chariker et al., 2011, for another example of applying basic learning principles to a spatial domain). It is also notable that in our spatial task, generation involved drawing, which has been shown to have benefits for learning science (Ainsworth et al., 2011), rather than retrieval, as in typical studies of the generation effect. In our experiments, we examined two forms of feedback: verbal feedback provided by an experimenter (in Experiment 2), on the accuracy of students drawing, and model-based feedback discovered by the student in attempting to match models to the drawings they had produced (Experiment 1). We now review what we have learned about each of these learning principles in the present research.

First, students improved from pretest to posttest in all conditions of our experiments, including the control conditions, indicating that merely being asked to generate diagrams is somewhat effective for learning (an example of the generation effect). Drawing forces students to commit to the relative spatial positions of components of a diagram, making them aware that the components can take different positions, and this alone might be sufficient to reveal gaps in students' knowledge, so that they learn to preserve the spatial relations correctly. However, it should be noted that the pretest-to-posttest improvements in the control groups were relatively small in both experiments, so generation alone is not a very effective strategy for teaching this skill.

An important point made by our research is that in spatial domains such as chemistry, not all forms of feedback are equally effective. It matters how feedback is given. In Experiment 2, providing students with verbal feedback, including the nature of the error, did not improve performance relative to the control group. In contrast, in Experiment 1, model-based feedback led to large, significant improvements in performance relative to the control group. A likely explanation of this result is that receiving model-based feedback also suggested a strategy for performing the task.

Although concrete models have the potential to be effective spatial tools, the evidence of their effectiveness has been mixed (e.g., Garg et al., 1999; Kaminski et al., 2009; McNeil & Uttal, 2009), and researchers have called for the need to develop pedagogical methods for how to best use models in instruction. In documenting the effectiveness of model-based feedback, our research suggests a novel and effective use of models in instruction. We propose that models present an interesting opportunity to generate self-feedback, that is, feedback that is discovered by the student and grounded in reality, rather than being provided by a tutor. Although self-feedback is not a new notion, it has received relatively little attention in the empirical literature (Hattie & Timperley, 2007). Possible advantages of self-feedback using a model are that it is more compelling, less humiliating (threatening) than feedback provided by another person (such as a teacher or peer), it helps students to develop self-regulatory mechanisms, and it helps students to appreciate the benefits of concrete models.

Although we examined the use of models in an organic chemistry, the challenges of developing representational competence and, particularly, reasoning about 3-D spatial relations from 2-D diagrams are evident in other domains, including anatomy, astonomy, geology, and geometry (Chariker et al., 2011; Cohen & Hegarty, 2014; Kali & Orion, 1996; Padalkar & Ramadas, 2011). The general approach here, of first generating diagrams and then receiving model-based feedback, can and indeed is already beginning to be applied in other domains. For example, Cohen and Hegarty (2014) had students use an interactive computer visualization to generate self-feedback in a task in which they had to imagine and draw cross-sections of simple solids and found that using this spatial tool for self-feedback was highly effective A similar approach is being used to develop penetrative thinking in geology (Gagnier, Atit, Ormand, & Shipley, 2012). In sum, there is evidence that generating diagrams followed by model-based feedback is effective for improving representational competence in a variety of domains.

Improved performance on the posttests in both studies was mediated by model use, raising the concern that models might have become a crutch to students, rather than scaffolding their learning. Therefore, it is important to test whether the effects of model-based feedback can generalize to situations in which students do not have access to models. Further studies are also necessary to examine the durability of the learning effects observed here. In ongoing research, we are addressing both of these issues, and the results are promising. Specifically, in a recent study, students who went through a similar intervention to that in Experiment 1 were more accurate on a posttest than a control group who received only verbal feedback, and this effect was still evident when they performed diagram translation problems without models available and after a 7-day delay (Stull & Hegarty, 2014). This suggests that our intervention can have lasting effects and that models can act as a scaffold to learning, and not just as a crutch. The fact that experiencing the benefits of models improved students' attitudes towards models, and not just the use of models in our specific task, also suggests that our interventions are likely to have long-term effects on students' learning.

In summary, we provided evidence for the effectiveness of a novel use of models in chemistry instruction in which students first generate solutions (diagrams) to problems and then use models to generate self-feedback on their solutions. The most effective intervention (in Experiment 1) took a short time (an average of 17 min to check the six pretest problems) and could be accommodated in a laboratory or tutorial session in the context of an organic chemistry class. Moreover, it can easily be adopted and tested in other disciplines such as anatomy, geology, astronomy, and architecture in which students must develop an understanding of complex 3-D structures.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16,* 183–198. doi:10.1016/j.learninstruc.2006.03.001

Ainsworth, S., Prain, V., & Tyler, R. (2011). Drawing to learn in science. *Science, 333,* 1096–1097. doi:10.1126/science.1204153

American Chemical Society. (2010). *2010 First Term Organic Exam* Retrieved from http://chemexams.chem.iastate.edu/materials/details_exam.cfm?ID=41754

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182. doi:10.1037/0022-3514.51.6.1173

Bjork, E. L., DeWinstanley, P. A., & Storm, B. C. (2007). Learning how to learn: Can experiencing the outcome of different encoding strategies enhance subsequent encoding? *Psychonomic Bulletin & Review, 14,* 207–211. doi:10.3758/BF03194053

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XV11* (pp. 435–459). Cambridge, MA: MIT Press.

Brown, A. L., Campione, J. C., & Day, J. D. (1981). Learning to learn: On training students to learn from texts. *Educational Researcher, 10,* 14–21. doi:10.3102/0013189X010002014

Bruice, P. Y. (2010). *Organic chemistry* (6th ed.). New York, NY: Pearson.

Chariker, J. H., Naaz, F., & Pani, J. R. (2011). Computer-based learning of neuroanatomy: A longitudinal study of learning, transfer, and retention. *Journal of Educational Psychology, 103,* 19–31. doi:10.1037/a0021680

Cohen, C. A., & Hegarty, M. (2014). Visualizing cross sections: Training spatial thinking using interactive animations and virtual objects. *Learning and Individual Differences, 33,* 63–71.

de Winstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32,* 945–955. doi:10.3758/BF03196872

diSessa, A. A. (2004). Metarepresentaion: Native competence and targets for instruction. *Cognition and Instruction, 22,* 293–331. doi:10.1207/s1532690xci2203_2

Downs, R. M., & Liben, L. S. (1991). The development of expertise in geography: A cognitive-developmental approach to geographic education. *Annals of the Association of American Geographers, 81,* 304–327. doi:10.1111/j.1467-8306.1991.tb01692.x

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions on Psychological Science, 12,* 83–87. doi:10.1111/1467-8721.01235

Gagnier, K. M., Atit, K., Ormand, C., & Shipley, T. F. (2012). *Improving penetrative thinking via progressive alignment and directed sketching*. Talk presented at The Geological Society of America Annual Conference, Charlotte, NC.

Garg, A., Norman, G. R., Spero, L., & Maheshwari, P. (1999). Do virtual computer models hinder anatomy learning? *Academic Medicine, 74,* S87–S89. doi:10.1097/00001888-199910000-00049

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77,* 81–112. doi:10.3102/003465430298487

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences, 8,* 280–285. doi:10.1016/j.tics.2004.04.001

Hinze, S. R., Rapp, D., Williamson, V. M., Shultz, M. J., Deslongchamps, G., & Williamson, K. C. (2013). Beyond ball-and-stick: Students' processing of novel STEM visualizations. *Learning and Instruction, 26,* 12–21. doi:10.1016/j.learninstruc.2012.12.002

Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 484–494. doi:10.1037/0278-7393.14.3.484

Kali, Y., & Orion, N. (1996). Spatial abilities of high-school students in the perception of geologic structures. *Journal of Research in Science Teaching, 33,* 369–391. doi:10.1002/(SICI)1098-2736(199604)33:4<369::AID-TEA2>3.0.CO;2-Q

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. (2009). Transfer of mathematical knowledge: The portability of generic instantiations. *Child Development Perspectives, 3,* 151–155. doi:10.1111/j.1750-8606.2009.00096.x

Keehner, M., Hegarty, M., Cohen, C. A., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science, 32,* 1099–1132. doi:10.1080/03640210801898177

Keig, P. F., & Rubba, P. A. (1993). Translation of representations of structure of matter and its relationship to reasoning, gender, spatial reasoning and scientific prior knowledge. *Journal of Research in Science Teaching, 30,* 883–903. doi:10.1002/tea.3660300807

Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 106–114. doi:10.1037/a0033699

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989–998. doi:10.1037/a0015729

Kozma, R., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of Research in Science Teaching, 43,* 949–968. doi:10.1002/(SICI)1098-2736(199711)34:9<949::AID-TEA7>3.0.CO;2-U

Kozma, R., & Russell, J. (2005). Multimedia learning of chemistry. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 409–428). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511816819.027

Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11,* 65–100. doi:10.1111/j.1551-6708.1987.tb00863.x

Lemke, J. L. (1998). Multiplying meaning. In J. R. Martin & R. Veel (Eds.), *Reading science: Critical and functional perspectives on discourses of science* (pp. 87–113). London, United Kingdom: Routledge.

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science, 29,* 587–625. doi:10.1207/s15516709cog0000_15

McNeil, N. M., & Uttal, D. H. (2009). Rethinking the use of concrete materials in learning: Perspectives from development and education. *Child Development Perspectives, 3,* 137–139. doi:10.1111/j.1750-8606.2009.00093.x

National Research Council, National Committee on Science Education Standards and Assessment. (1996). *National science education standards*. Washington DC: National Academies Press.

Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a Gestalt perceptual principle. *Journal of Experimental Psychology: Applied, 13,* 197–223. doi:10.1037/1076-898X.13.4.197

Padalkar, S., & Hegarty, M. (2013, February 7–11). Undergraduate students' understanding about representations in chemistry. In G. Nagarjuna, A. Jamakhandi, E. B. Sam (Eds.), *Proceedings of Conference episteme-5* (pp. 288–294). Mumbai, India.

Padalkar, S., & Ramadas, J. (2011). Designed and spontaneous gestures in elementary astronomy education. *International Journal of Science Education, 33,* 1703–1739. doi:10.1080/09500693.2010.520348

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26,* 521–562. doi:10.1207/s15516709cog2605_1

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16,* 475–522. doi:10.1207/s1532690xci1604_4

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22,* 129–184. doi:10.1207/s1532690xci2202_1

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153–189. doi:10.3102/0034654307313795
Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604. doi:10.1037/0278-7393.4.6.592
Stieff, M., Scopelitis, S., Lira, M., & DeSutter, D. (2013). *Improving representational competence in organic chemistry with concrete models*. Manuscript submitted for publication.
Stull, A., & Hegarty, M. (2014, April). *Developing representational competence in chemistry with concrete and virtual models*. Paper presented at the annual conference of the American Educational Research Association, Philadelphia, PA.
Stull, A. T., Hegarty, M., Dixon, B., & Stieff, M. (2012). Representational translation with concrete models in organic chemistry. *Cognition and Instruction, 30,* 404–434. doi:10.1080/07370008.2012.719956
Treagust, D. F., & Chittleborough, G. (2001). Chemistry: A matter of understanding representations. In J. Brophy (Ed.), *Subject-specific instructional methods and activities* (Vol. 8, pp. 239–267). New York, NY: Elsevier.

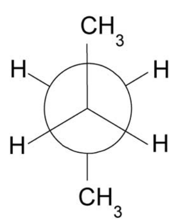
Appendix

Instruction Sheet Provided to the Participants in the Beginning of the Experiments

Instructions

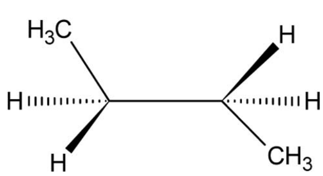
Welcome and thanks for agreeing to participate in our study! This study is about diagrams and models in organic chemistry. In this study, you will complete 18 worksheets. For each of the worksheets you will be given a Newman, Dash-Wedge, or Fischer projection (diagram) of a molecule as well as a physical model of that molecule. Your task is to draw a different projection for each molecule. For example, you might be given a dash-wedge projection of a molecule and asked to draw the corresponding Newman projection for the same molecule. The text at the top of each page will describe which projection you are to draw. Some of the transformations may be difficult but please try your best. Before we proceed to the worksheets, we will review the rules for interpreting each of the different projections that you will be expected to draw. Below are examples of the three projections. All three use different conventions to illustrate the 3-D shape of the molecule. The same 4-carbon molecule is illustrated in all three projections in the examples below.

Newman



In a Newman projection, the molecule is oriented with one backbone carbon in front of the other. The front carbon is located at the intersection of the 3 lines (noon, 4 o'clock and 8 o'clock around the circle). The atoms at the ends of these three lines are attached to the front carbon. The rear carbon is behind the circle. The atoms at the ends of the shorter lines connected to the circle (2 o'clock, 6 o'clock, and 10 o'clock around the circle) are attached to the rear carbon.

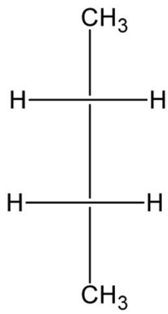
Dash-Wedge



In a Dash-Wedge projection, the molecule is oriented with the backbone carbons at the two 4-way intersections of lines on the left and right of the diagram. Dashed lines represent bonds to atoms that are going into the page (below the plane of the paper). Wedge lines represent atoms that are coming out of the page (above the plane of the paper). Solid lines represent bonds to atoms that are on the plane of the paper.

(*Appendix continues*)

Fischer



The Fischer projection to the left illustrates a 4-carbon molecule. The atoms at the right and left of the horizontal lines are coming out of the page (above the plane of the paper) and the atoms at the top and bottom of the vertical line are going into the page (below the plane of the paper). The two backbone carbons are located where the horizontal lines cross the vertical line. These carbons are on the plane of the paper.

Take a moment to visualize how each projection represents the three-dimensional structure of the molecule and satisfy yourself that the three above projections represent the same molecule. Compare and contrast the three projections because you will need to draw each in the following activity.

Please let the experimenter know if you have any questions.

Received September 6, 2013
Revision received May 12, 2014
Accepted May 18, 2014 ■